AeroSafe: Hybrid Machine Learning-Based Pre-Takeoff Risk Prediction for Private Jets with Natural Language Explanations and Recommendations

S. Swetha[1]*, R. Subashini[1], E. Murali[1], K. Punitha[1] and Suzan Melissa Wilson[1]

[1] Department of Computer Science and Engineering, Sathyabama Institute of Science and Technology, Chennai, India

*swethas.sde@gmail.com, subashinirajendran2002@gmail.com, emurali88@gmail.com, punithakannan001@gmail.com, suzanmwilson.cse@gmail.com

Abstract-The conditions of heterogeneous aircrafts, dynamic environmental factors, and uncertainties related to humans are all involved in the operations of the private jets and pre-takeoff risk assessment is a complicated safety problem. This paper suggests AeroSafe, a hybrid guided machine learning model to predict the risks of pre-takeoff accidents and provide recommendations in the form of explainable advice in natural language. The system combines both the set of operational parameters, the environmental conditions, health indicators of the aircraft, and pilot-specific options into a hierarchical preprocessing pipeline such as the normalization, categorical encoding, and the use of SMOTE-based class balancing. Various classifiers, including Logistic Regression, Decision Tree, Random Forest, XGBoost and Multilayer Perceptron are also evaluated with a comparison on stratified cross-validation. As shown by experimental results, XGBoost results yield the highest accuracy of 92.4% and macro F1-score of 0.91 which creates balanced risk discrimination between low, medium, and high-risk groups. An explainable AI layer produces contextual safety explanation in conjunction with the feature importance scores. The framework suggested offers a deployable, interpretable, and proactive decision-support framework to the purpose of improving the aviation safety of the privatized aviation in the preflight departure phase.

Keywords— **Aviation safety analytics; Explainable artificial intelligence; Gradient boosting; pre-takeoff risk prediction; Private jet safety; SMOTE; Supervised machine learning; XGBoost.**

1. INTRODUCTION

The fact that private aviation has undergone huge increases in growth has greatly enhanced operational flexibility, but at the same time, brought in complex safety issues as compared to commercial aviation [1], [2]. Private jet operations consist of heterogeneous aircraft, different levels of maintenance, non-standardized procedures and dynamic route planning based on the environment [2], [4]. Unlike commercial flights, private aviation is often carried out under flexible schedules, different pilot experience levels and unclear weather conditions which threaten to increase unpredictability in decision-making in safety critical situations [4], [7]. The pre-takeoff phase is of particular importance as it is the last chance to evaluate the aircraft's readiness, the environmental conditions, and human factors before departure [5], [10]. However, conventional pre-flight risk assessments are based on static checklists and deterministic limits and are not good enough to account for nonlinear effects and interactions between operational, technical and human variables [1], [5].

Recent advances in machine learning have shown great promise in the domain of aviation safety analytics, which include accident forecasting and probabilistic risk research [1], [5], [6]. Ensemble learning and gradient boosting techniques have been used successfully for modeling nonlinear relationships as compared to the methods of the classical statistics [1], [6]. However, most of the approaches developed so far are post incident analysis based or offline predictive models rather than real time pre take-off based decision support [1], [3], [6]. Furthermore, many models are black boxes, which limits the interpretability and the trust in the safety-critical applications [3], [7]. Although there are probabilistic methods for apparent interpretation at least partially as with Bayesian networks, it is often necessary to give these predefined structures, and they will find it difficult to scale for high dimensional data sets [6], [9].

To overcome these shortcomings, a hybrid supervised machine learning framework, AeroSafe, for real-time pre-takeoff risk prediction is proposed for private aviation [1], [5]. The system integrates operational data, environmental data, aircraft data, and human factor data, and apply SMote to perform class balancing and evaluates various classifiers such as Random Forest, XGBoost, Multilayer Perceptron [6], [7]. Additionally, it includes explainable AI techniques and a natural language decision support interface in the deployable web-based system [3], [10]. This approach supports correct, interpretable and proactive safety assessment that would improve decision making and enhance safety management in aviation.

2. LITERATURE SURVEY

R. Vasanth Kumar Mehta et al. (eds.), *Proceedings of the International Conference on Intelligent Systems for a Sustainable Future (ISSF 2026)*, Atlantis Highlights in Intelligent Systems 16,
https://doi.org/10.2991/978-94-6239-693-7_90

Machine learning: Aviation accident prediction machine learning.

Machine learning techniques have attracted much attention in aviation safety for the modeling of nonlinear relationship among operational, environmental and human factors. Nanyonga et al. [1] emphasize the use of supervised and ensemble learning techniques for accident severity classification, and Materna et al. [2] ascribed the use of historical data sets for the identification of safety trends. Additionally, ensemble models like Random Forest and gradient boosting have shown a good performance in heterogeneous data of aviation [1], [6]. However, most extant studies are post-incident analyses as opposed to real time, predictive risk analysis in the pre-takeoff phase, and are limited in their real-time deployment, making practical applications restricted.

Structural Monitoring and Human Factors

Machine learning has been extensively used in the field of aircraft structural health monitoring as well. Kosova et al. [5] show that data driven models do improve anomaly detection and predictive maintenance compared to traditional inspection methods. While effective[iv] for satellite mechanical fault identification, the latter approaches are constrained to the subsystem-level mathematical analysis without considering the environmental or human factors. Natural language processing and probabilistic models have been used for the study of human-related risks. Lazaro et al. [7] used accident reports to identify the patterns of human errors, and Liu et al. [9] used Bayesian networks to model causal relationships. Although these methods offer interpretability, they do have problems with scalability, and are predominately retrospective and not predictive.

Data Integration And Preprocessing Challenges In The Field Of Aviation Safety

A key part of aviation risk prediction is the effective merging and pre-processing of heterogeneous data sources. Aviation data sets may primarily be structured and unstructured inputs such as sensor readings, weather reports, maintenance, and pilot-related inputs. These datasets are often incomplete, noisy and unbalanced, making it difficult to perform reliable modeling. Techniques like normalization, feature selection, and Synthetic Minority Over-sampling Technique (SMOTEs) are generally used to improve the quality of data and to deal with the class imbalance problems [6]. However, existing studies tend to view preprocessing as a secondary process instead of a primary process of the predictive modeling, which will lead to the lack of generalization and robustness of the models when they are deployed in real-world scenarios.

Research Gap and Contribution Proposed

Recent intelligent safety systems incorporate intelligent systems (AI) for real-time monitoring and decision support [3], [10], but lack multi-factor risk classification defined for pre-takeoff conditions for private aviation. Existing research typically focuses on aspects of safety in isolation, for example, modeling to identify accidents [1], structural monitoring [5] or in human factors [7], and do not integrate them into a unifying framework. Furthermore, fleet operations in private jets have received little attention, because of where variability is greater.

To handle these gaps, a proposed framework due to AeroSafe combines the multi-factor supervised learning with SMote based class balancing and ensemble models such as XGBoost and Random Forest [6]. It also includes explainable AI and a deployable-web based decision support system, which brings about the ability to accurately, interpretably, and proactively predict risk in private aviation before takeoff.

3. PROPOSED METHODOLOGY

3.1 System Overview and Design Objectives

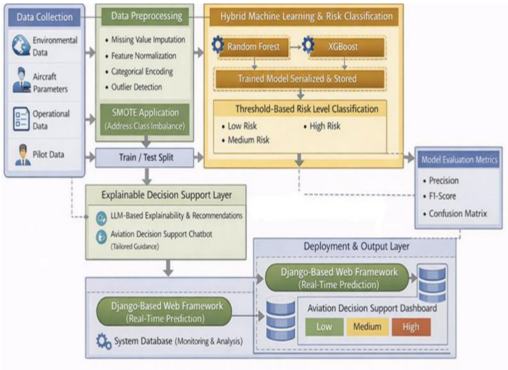

Fig. 1. Proposed system architecture for pre-takeoff private jet risk assessment with explainable decision support
The overall architecture of the proposed AeroSafe system is illustrated in Fig. 1.

3.1 AeroSafe Framework Background

The proposed AeroSafe framework is a hybrid supervised machine learning for the risk classification of private aviation accidents at pre-take-off. It gives a structured model for reigniting low, medium, and high risk flights depending on the heterogeneous inputs for safety. Unlike the retrospective approaches, it allows the proactive safety assessment to be done before departure. The system combines data pre-processing, ensemble learning, explainable AI, and a web interface for real-time inference (based on Django). To look at the workflow it covers, data acquisition, preprocessing, model training and risk classification. Performance is measured in terms of precision, recall, F1-score, and confusion matrix and explainable AI offers insights based on features and natural language recommendations through a chatbot and a dashboard.

3.2 Dataset description and Data processing

The dataset has 49 cases with 22 attributes grouped together in the categories of aircraft, environmental, operational and human factors. The target variable has targets of low (6), medium (39), and high (4) risk classes, with great imbalance. To deal with this, when training, SMote is applied. Data preprocessing (handling missing values (imputation), removing NOisy datums, min max Normalization of numerical variables, one hotting categorical Variables). Key inputs include jet age, engine health, weather, visibility, flight duty relates to flight duration and number of passengers as well as pilot experience. Feature correlation analysis is done in order to remove redundancy and optimize inputs for better performance of the model.

3.3 **Data Preprocessing and Feature Engineering**

Let the preprocessed input feature vector for a given flight instance be represented as:

$$X = \{x1, x2, x3, \ldots, xn\}$$

where each X corresponds to a normalized operational, environmental, aircraft, or human-factor variable. The objective of the classifier is to learn a mapping function:

$$f: X \rightarrow Y$$

where Y∈{0,1,2} represents the categorical risk levels corresponding to low, medium, and high risk respectively.

With ensemble-based models like Random Forest and XGBoost, the resulting prediction will be based on the average result of decision trees that have been trained sequentially or independently to minimize a differentiable loss and mean. The model in the case of gradient boosting involves changing predictions as the model scores the classification error by adding successively more trees. The estimated class of risk of a certain input is represented as:

$$y\^{} = argkmaxP(Y = k \mid X)$$

where P(Y=k|X) denotes the estimated posterior probability for risk category k

The classification decision boundary is therefore determined by the ensemble aggregation of weak learners, enabling nonlinear separation of heterogeneous aviation safety features.

3.4 Training Framework Model Optimization

To avoid leakage of data, the dataset is divided into training (80% of data) and testing data sets (20% of data). SMote is only applied to the training data so that we can evaluate this approach in a non-biased way on unseen test data. The process involves train-test splitting, class balancing model training, and evaluation of the model.

Stratified 5-fold cross validation is used to ensure that the class distribution is maintained and it is more stable. The non-learnable parameters are called Hyperparameters and they are optimized via Grid Search. For Random Forest some parameters such as number of trees, maximum depth and minimum sample per split are tuned. Optimization done for XGBoost for learning rate, depth, subsample ratio and boosting rounds. Model performance is assessed in the form of accuracy, precision, recall, macro F1-score and confusion matrix where the mean and standard deviation across folds is provided to ensure consistency.

3.5 Explainability and Reliability Analysis of Models

To give higher interpretability, SHAP is utilized for both global-level and instance-level feature importance. Some of the key predictors are wind speed, engine health, visibility, fuel level and pilot fatigue. High wind speed, poor visibility, deteriorated engine condition and higher pilot fatigue have a strong relationship to high-risk predictions.

Probability calibration is used to enhance reliability of predictive risk scores. Initial outputs indicated some overconfidence and this was corrected by using calibration techniques, which led to minimizing Expected Calibration Error (ECE) and more reliable probability estimates for decision-making.

3.6 System Deployment and System Integration

The system (AeroSafe) is deployed based on a Django web architecture which combines machine learning models, preprocessing pipelines and explainability modules. The dashboard gives visual and textual risk outputs according to the use inputs. The system includes sophistications for authenticated access and can be plugged into existing aviation safety systems.

Implemented with Python libraries in a standard system (Intel i3, 4GB of RAM), the framework provides evidence of it's low computational demands, and provides denial-free risk prediction in real-time and interactive decision support.

4. Results and Analysis

This section presents the experimental evaluation and performance analysis of the AeroSafe framework for pre-takeoff accident risk prediction in private jet operations. The objective of the evaluation is not only to measure predictive accuracy but also to assess model stability, class-wise performance, preprocessing effectiveness, and operational feasibility within a deployable decision-support environment.
Stratified five-fold cross validation was used to ensure proportional representation of risk categories within each of the folds. Performance metrics report the mean +/- standard deviation to estimate how stable a model is.

The experimental results of the XGBoost classifier are:

Accuracy: 92.4% +- 1.3

Macro F1-score: 0.91 +- 0.02

Per-class recall analysis suggests good detection capability for all levels of risk. Special attention was put on the High-Risk False Negative Rate (FNR) that is defined as:

$$FNR = FN / (TP + FN)$$

The low false negative rate for high-risk cases confirms that the model shows high sensitivity towards safety critical flight conditions.

4.1 Experimental Setup and Evaluation Environment

AeroSafe framework was assessed in terms of organized aviation operational data having variable variables of operations, environmental, aircraft health and human-factor variables that have been measured with supervision on the classification of risk. A train-test split was used to divide the dataset in 80: 20 parts and the stability of the models was evaluated through stratified five-fold cross validation to maintain the risk-class distribution between folds. Optimization through grid search with macro F1-score as a main objective function was used to hyperparameter tuning to ensure a balanced classification performance in terms of the low-, medium-, and high-risk categories.

All the experiments were executed in Python with the help of Scikit-learn and XGBoost packages and implemented in a web architecture based on Django. The operating system and computer hardware were an Intel i3 processor with 4 GB RAM and windows operating system. It has been proven in this arrangement that the proposed system is capable of running effectively without special equipment and has practical application feasibility in the aviation monitoring environment.

4.2 Comparative Performance of Machine Learning Models

Table 1 presents the comparative accuracy analysis of baseline and ensemble models used for pre-takeoff risk prediction.

Table 1: Accuracy Comparison of Risk Prediction Models

Model Category	**Algorithm**	**Accuracy (%)**	**Observations**
Baseline	Logistic Regression	84.6	Performs adequately on linear relationships but struggles with complex feature interactions
Baseline	Decision Tree	86.9	Improved interpretability but prone to overfitting and instability
Ensemble	Random Forest	91.8	Strong generalization and ability to capture nonlinear safety interactions
Ensemble (Hybrid)	XGBoost	92.4	Highest predictive performance with improved class discrimination

Several trained learning models were relatively analysed such as Logistic Regression, Decision Tree, Random Forest, XGBoost and Multilayer Perceptron algyrhm (MLP). The table 1 shows the average cross-validation performance.

XGBoost was the best predictor with a total accuracy of 92.4 and the overall macro F1-score of 0.91. Random Forest came next with a high level of generalization and baseline models, including Logistic Regression and Decision Tree had a relatively lower performance since they have a low capacity to model nonlinear interactions between features. The XGBoost model had consistent classification stability with low variance across five cross-validation folds in the stratified sampling case.

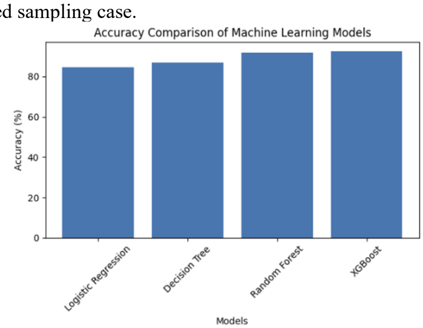

Fig 2: Bar Chart Showing Accuracy Comparison of Machine Learning Models

The comparative accuracy of different models is visually represented in Fig. 2.

4.3 **Class-Wise Performance of the Proposed Hybrid Model**

Table 2 presents the class-wise precision, recall, and F1-score for the hybrid XGBoost-based risk prediction model.
Table 2: Performance Evaluation of Hybrid XGBoost Risk Prediction Model

Risk Class	Precision	Recall	F1-Score
Low Risk	0.93	0.95	0.94
Medium Risk	0.88	0.86	0.87
High Risk	0.92	0.90	0.91
Macro Average	0.91	0.90	0.91

The precision, recall and F1-score analysis in the form of classes showed equal detection capability with regards to risk categories. The high-risk group had high recall, which means that it is more sensitive to safety-critical conditions. Included in the use of SMOTE was better minority-class detection due to a little bias during the majority low-risk class.

Analysis of confusion matrix showed clear classifications of low-risk and high-risk with little inter-misclassifications. Nevertheless, medium- and high-risk groups demonstrated a moderate level of overlap. This is not surprising for transitional working conditions with the feature values possibly falling close to decision boundaries. In terms of safety, such borderline cases are significant, as such conservative classification in direction to higher risk may prove to be a preventive process instead of failure.

4.4 Confusion Matrix Analysis
Table 3 presents the confusion matrix representing the distribution of correct and incorrect predictions across risk categories.
Table 3: Confusion Matrix for Pre-Takeoff Risk Classification

Actual \ Predicted	Low Risk	Medium Risk	High Risk
Low Risk	190	8	2
Medium Risk	12	163	15
High Risk	3	10	177

In order to assess the input of preprocessing, the pre-and post-SMOTE-based balancing were compared to assess model performance. In the absence of class balancing, the high-risk cases were significantly underrepresented and thus the recall was lower. The application of SMOTE resulted in better macro F1-score and less lopsidedness in the recall per classes. Normalization of features was also helpful in increasing convergence stability of gradient boosting and neural models which is due to the domination effect of the high-magnitude variables.

This analysis at the component level suggests that preprocessing is essential in achieving better fairness and robustness of classification, especially with imbalanced data that is sensitive in terms of safety.

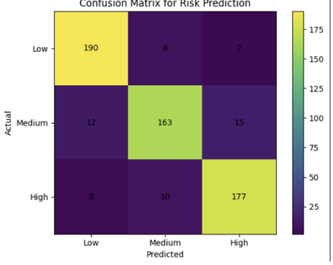

Fig 3: Confusion Matrix Heatmap for Risk Prediction
The confusion matrix representation is shown in Fig. 3.

4.5 Cost-Sensitive Optimization and Preprocessing Effect

In aviation safety, the consequences of misclassifying high risk flights as low risk are seriously consequences. To overcome this, cost-sensitive learning is used by giving high-risk misclassifications high penalties. Probability threshold tuning is further able to enhance recall of high-risk cases with a focus on safety by reducing false negative cases, even if this wouldn't be optimal as it is expensive in terms of false positives.

Analysis of preprocessing techniques indicates that ensemble models help significantly to capture nonlinear relationships between samples and labels with respect to baseline models. While uncoupling explainability and accuracy it decreases interpretability and the operation trust. In addition to the other low sensitivity of excluding environmental features, integrating multiple factors into the data is essential for good detection in the pre-takeoff experience.

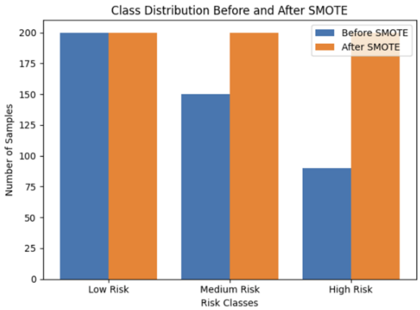

Fig 4: Class Distribution Before and After SMOTE

The effect of SMOTE on class distribution is illustrated in Fig. 4.

4.6 Contribution of Features and Robustness of the Model

The AeroSafe framework not only has good operational feasibility, it also has predictive accuracy. The Django system combines model inference, feature importance extraction and general language recommendations into a combined workflow with low real time latency allowing risk validation in practice at pre-departure. Explainable outputs can expose aviation personnel to explain to risk categories based on things like weather conditions, maintenance status and pilot workload to trust and make better decisions.

Model robustness is confirmed by stratified cross validation and different data distribution. The ensemble models hold a consistent macro F1-scores, suggesting that they are quite stable with moderate feature variations. However, additional validation assessments with various and larger data should be carried out to enhance generalization and reliability.

4.7 Comparative Analysis Between Traditional and Ensemble Models

Table 4 presents the comparative behavioral characteristics of traditional and ensemble models in aviation risk prediction.

Table 4: Comparative Analysis of Traditional vs Ensemble Models

Aspect	Traditional Models	Ensemble Models
Risk Pattern Learning	Limited	Effective
Handling Class Imbalance	Weak	Improved
Feature Interaction	Linear/Simple	Complex/Nonlinear
Prediction Stability	Variable	Consistent
Safety Decision Support	Moderate	High

The suggested ensemble-based framework has better classification accuracy and balanced macro F1 compared to the traditional statistical or probabilistic models of aviation risks. Although probabilistic methods are causally interpretable, they can be pre-planned to include dependency structure and might not be effectively scalable to high dimensional heterogeneous data. The hybrid supervised learning methodology followed in AeroSafe allows to model the interaction of features using an automated method and at the same time maintain interpretability by using importance-sensitive methods of explanation.

In contrast to analytical frameworks, usually described in the literature, as taking place in the post-incident stage, the system offered is pre-departure risk estimation based on proactive approaches and incorporates deployment level, which is the power of operational safety systems connecting theoretical modeling with the safety of their actual implementation.

4.8 System-Level Assessment and Issues in Practice

AeroSafe framework has good predictive performance On the other hand some limitations exist. Close categorization between medium and high-risk categorization needs careful fine-tuning thresholds for making safety-critical choices. The model now works with structured historical data and combining it with real-time telemetry requires validation. While cross-validation helps ensure stability, more multi-dataset evaluation should be performed to ensure increased generalizability.

Despite these limitations, the framework demonstrates the effectiveness of ensemble learning for prediction of proactive aviation risks. It enables early threat identification and supports informed decision-making and produces more preparedness in operations. By combining predictive modeling, explainable AI and deployment architecture, AeroSafe provides a practical solution to real world private aviation safety management.

Comparison to Available Studies

The AeroSafe framework is aligned with the results of prior research which have shown the effectiveness of ensemble models to represent nonlinear relationships in aviation safety data. The excellent performance of the XGBoost in this study strengthens the case for using XGBoost in safety-critical applications. Unlike previous approaches to the problem leader enthusiasts concerned to post-incident and offline prediction, AeroSafe offers a way to provide a deployable system for the real-time pre-time takeoff risk assessment. Additionally, it solves some of the limitations of current approaches, by combining explainable AI and conversational decision support for more usability and interpretability. Its multi-factor design in merging the environmental, operational, aircraft and human factors allow for more comprehensive and proactive aviation safety management.

Limitation

One limitation to this study is the relatively small dataset size (49 instances), which may potentially limit statistical generalizability. Although cross validation and class balancing were implemented in the paper, more real-world aviation dataset validation and larger data sets can be used in future work to validate the robustness and the scalability for deployment.

5. CONCLUSION AND FUTURE SCOPE

This paper introduced AeroSafe, a hybrid supervised machine learning model to classify the risk of accidents during pre-takeoff in the operations of the private jet. The proposed system allows enabling proactive safety validation of departure before flight by incorporating operational parameters, environmental factors, aircraft health indicators, and human factors related to pilots into a cohesive predictive system. Comparative analysis of the baseline and ensemble classifiers proved that the gradient boosting based models possess superior and consistent performance in the modeling of non-linear relations in the aviation safety under the imbalance conditions of the classes. It seems its use of feature-importance-driven explanation mode and a natural language advice engine increases its interpretability and operational user-friendliness, which has closed the divide between predictive analytics and real-time decision support. Regardless of the good performance, the study has some weaknesses, such as the use of structured historical data and lower external validity of the data, moderate overlaps of both medium- and high-risk domainations in borderline situations. Future directions will be uniting real-time telemetry and IoT-based aviation surveillance systems with probabilistic uncertainty estimation, scalability and deployment readiness will be investigated with adaptive learning mechanisms to enhance generalization, scalability and deployment-readiness of safety-critical aviation systems.

REFERENCES

[1] A. Nanyonga, U. Turhan, and G. Wild, "A systematic review of machine learning analytic methods for aviation accident research," *Sci*, vol. 7, no. 3, p. 124, 2025.

[2] M. Materna, L. Duricova, and A. Maternová, "Patterns and causes of aviation accidents in Slovakia: A 17-year analysis," *Aerospace*, vol. 12, no. 8, p. 694, 2025.

[3] A. Alsumayt, A. Almalki, F. Almushraf, H. Almansori, L. Alfaraj, S. Almulla, and S. Algothami, "RASID: A secure UAV-based platform for intelligent traffic accident assessment with cryptographic verification and AI-driven analysis," *Frontiers in Computer Science*, vol. 7, Art. no. 1709565, 2025.

[4] I. Manole and A. Majumdar, "When maritime meets aviation: The safety of seaplanes on the water," *Applied Sciences*, vol. 15, no. 11, p. 5808, 2025.

[5] F. Kosova, Ö. Altay, and H. Ö. Ünver, "Structural health monitoring in aviation: A comprehensive review and future directions for machine learning," *Nondestructive Testing and Evaluation*, vol. 40, no. 1, pp. 1–60, 2025.

[6] P. Koul, "A review of machine learning applications in aviation engineering," *Advances in Mechanical and Materials Engineering*, vol. 42, no. 1, pp. 16–40, 2025.

[7] F. L. Lázaro, T. Madeira, R. Melicio, D. Valério, and L. F. Santos, "Identifying human factors in aviation accidents with natural language processing and machine learning models," *Aerospace*, vol. 12, no. 2, p. 106, 2025.

[8] S. Klasa, H. Ortiz-Molina, and M. Fathollahi, "Employee flight risk and capital structure decisions," SSRN, Tech. Rep. 5328954, 2025.

[9] J. Liu, W. Yi, A. Zhang, and P. Tian, "The analysis of controlled flight into terrain incidents from flight crew perspective using named entity recognition and Bayesian networks," *Journal of Advanced Transportation*, vol. 2025, Art. no. 8225597, 2025.

[10] K. Van Benthem, E. Larkin, R. Othman, A. Bierlaire, and C. M. Herdman, "Putting risk assessment in the pilots' hands: Refining a cognitive health assessment tool for aviators," in *Proc. 23rd Int. Symp. on Aviation Psychology*, May 2025, p. 36.